HathiTrust

HathiTrust es una asociación de instituciones académicas y de investigación que ofrece una colección de millones de documentos digitalizados provenientes de bibliotecas de todo el mundo, y trabaja para asegurar la accesibilidad y preservación a largo plazo de los sus registros culturales.

«Hathi» significa «elefante»[1] en hindi y urdu, un animal célebre por su gran memoria. Su preservación es uno de los grandes activos de HathiTrust.

Índice

Introducción

Comunidad de socios

Misión y metas
- Misión
- Metas

Objetivos funcionales
- Objetivos a corto plazo
- Objetivos a largo plazo

La biblioteca digital

Políticas
- Acceso y uso
- Corrección de metadatos bibliográficos
- Derechos de autor
- Eliminación de HathiTrust
- Preservación digital
- Directrices de los depósitos digitales
- Calidad
- Privacidad

Tecnología, estándares y certificaciones
- Aspectos tecnológicos
- Estándares de la biblioteca y el contenido digital
- Especificaciones para los objetos digitales
  - Definiciones
  - Especificaciones

El centro de investigación

Conclusiones

HathiTrust

Tipo	Biblioteca digital
Fundación	2008
Sede	Míchigan
Sitio web	www.hathitrust.org (https://www.hathitrust.org/)

Véase también
Referencias
Enlaces externos

Introducción

HathiTrust surgió en 2008 como una colaboración de las universidades del Comité sobre Cooperación Institucional (https://www.cic.net/about-cic) (*Committee on Institutional Cooperation*, CIC) y el sistema universitario público de la Universidad de California (*University of California*, UC), con el fin de crear un repositorio para archivar y compartir sus respectivas colecciones digitalizadas. HathiTrust se ha expandido con rapidez, incluyendo nuevos socios y proporcionándoles una manera simple de archivar su contenido digital.

El propósito original de la asociación fue preservar y proporcionar acceso al contenido digitalizado de libros y revistas pertenecientes a las colecciones de sus socios, incluido material con derechos de autor y de dominio público digitalizado por Google, el *Internet Archive*, Microsoft y otras iniciativas privadas. Los socios pretenden construir un archivo que abarque documentación publicada en todo el mundo, y establecer estrategias comunes para la gestión y el desarrollo colaborativo de su material digital e impreso.

La principal comunidad a la que HathiTrust presta servicio es la formada por los miembros (profesorado, estudiantes y usuarios) de sus bibliotecas asociadas, aunque su material es considerado un bien público a disposición de usuarios de todo el mundo, siempre y cuando la normativa contenida en leyes y contratos lo permita.

Comunidad de socios

HathiTrust es una comunidad internacional de bibliotecas académicas y de investigación consagradas al acceso y preservación a largo plazo de sus fondos culturales digitalizados. Mediante un esfuerzo común y haciendo gala de un profundo compromiso con el bien público, las bibliotecas respaldan actividades de enseñanza y aprendizaje del profesorado, los estudiantes o los investigadores en sus respectivas instituciones, así como las necesidades académicas del público en general.

En la actualidad HathiTrust cuenta con 116 socios,[2] de los cuales 4 son consorcios o sistemas estatales estadounidenses (el Comité de Cooperación Institucional, el Sistema Universitario Estatal de Florida, la Universidad de California y el Sistema Universitario de Texas), y el resto son instituciones individuales, incluida la Biblioteca del Congreso de Estados Unidos (*Library of Congress*, LoC).

En noviembre de 2010, la Universidad Complutense de Madrid pasó a ser socio de HathiTrust, y se convirtió en la primera universidad europea en formar parte de la comunidad.[3]

Misión y metas

Misión

La misión de HathiTrust es contribuir a la investigación, el estudio y el bien común recogiendo, organizando, preservando, comunicando y compartiendo de forma colaborativa la **memoria del conocimiento humano**.

Metas

Más específicamente, HathiTrust se encarga de:

- Construir un **archivo digital** de material bibliotecario -obtenido a partir de la conversión de las colecciones impresas de las instituciones que la integran- que sea fiable, y cuya propiedad y administración está cada vez compartida.
- Mejorar de manera drástica el **acceso** a este material, de manera que se satisfagan principalmente las necesidades de las instituciones copropietarias, y poniendo especial énfasis en garantizar el acceso a las personas con problemas de lectura.
- Desarrollar una **infraestructura** de contenidos digitales que sea rentable y sólida y aporte valor a estudiosos e investigadores, incluyendo gran variedad de formatos, además de materiales originalmente digitales.
- Definir asociaciones y **servicios** que garanticen la conservación de los materiales de HathiTrust y del conjunto de registros académicos impresos y digitales.
- Redoblar los esfuerzos de coordinación de estrategias de **almacenamiento** compartido entre bibliotecas para reducir los costos operacionales y de capital a largo plazo destinados al almacenamiento y cuidado de colecciones impresas.
- Construir una infraestructura que facilite la colaboración rentable y productiva entre las **instituciones asociadas**, para reducir el costo destinado a asegurar los activos intelectuales del campus.
- Definir y ofrecer un conjunto de servicios que hagan uso del corpus de HathiTrust para servir de apoyo a la **investigación**.
- Crear un **marco técnico** que permita la creación -tanto centralizada como distribuida- de herramientas y servicios.
- Considerar a la empresa HathiTrust como un «**bien público**», y a su vez definir un conjunto de servicios que beneficien a sus socios.

Objetivos funcionales

Objetivos a corto plazo

- **Mecanismo para visualizar y pasar páginas** (*Page Turner*). HathiTrust da soporte a una aplicación interactiva de lectura y descarga de textos e imágenes propias.
- **Marcado (iniciativa global; bibliotecas individuales)**. HathiTrust da soporte al marcado de documentos de sus fondos mediante etiquetas identificativas y marcas de agua de diverso tipo.
- **Validación, migración y comprobación de errores de formato**. Hasta la fecha no ha sido necesario realizar ningún tipo de migración, pero en previsión de esta, HathiTrust almacena diversos metadatos de preservación técnica y digital para cada documento. Además, se están desarrollando estrategias para asegurar y validar la integridad del material.
- **Desarrollo de APIs que permita a las instituciones asociadas acceder a la información e integrarse en los sistemas locales de manera individualizada**. HathiTrust dispone de una API bibliográfica que permite la búsqueda e integración en el catálogo, y una API de datos que ofrece acceso a los datos subyacentes de los recursos digitales.

- **Mecanismos de acceso para personas discapacitadas**. HathiTrust ha desplegado un interfaz accesible que facilita su navegación y uso por personas con discapacidades visuales.
- **Integración del catálogo de HathiTrust en WorldCat**. Se ha realizado un piloto de implementación del catálogo HathiTrust para incluirlo en WorldCat, el catálogo digital más grande del mundo, que forma parte de OCLC.
- **Capacidad para publicar colecciones virtuales**. HathiTrust ha creado una aplicación constructora de colecciones (*Collection Builder*) que permite a cualquier individuo crear colecciones públicas (compartidas) y privadas.
- **Mecanismo para incorporar de forma automática contenido no perteneciente a Google**. HathiTrust desarrolló un mecanismos de incorporación automática de contenido de libros y revistas digitalizados por el *Internet Archive*, y actualmente se está implementando un marco técnico y normativo para la incorporación del contenido de libros y revistas de otras instituciones asociadas.

Objetivos a largo plazo

- **Compatibilidad con los elementos necesarios de la lista de criterios y comprobación de la Auditoria y Certificación de Repositorios de Confianza** (*Trustworthy Repositories Audit & Certification*, TRAC). El Centro para Bibliotecas de Investigación está llevando a cabo una evaluación independiente del repositorio de HathiTrust, fundamentada en los criterios del TRAC.
- **Mecanismos robustos de descubrimiento, como búsqueda de texto completo a través de los repositorios**. Se ha publicado una implementación inicial de búsqueda de texto completo, que ha implicado un considerable esfuerzo en investigación y desarrollo y que ha sido ampliamente documentada.
- **Desarrollo de una definición de servicio abierto para hacer posible que las bibliotecas asociadas desarrollen otros mecanismos de acceso seguro y herramientas de descubrimiento**. HathiTrust ha creado diversas API con este fin, así como un entorno de desarrollo colaborativo para que los socios puedan realizar sus aportaciones y desarrollar nuevas aplicaciones.
- **Soporte de otros formatos más allá de los propios de libros y revistas**. HathiTrust está investigando y gestionando proyectos piloto sobre aspectos relacionados con el almacenamiento y entrega de publicaciones electrónicas (en particular, el formato .epub) y archivos de audio e imagen (como mapas).
- **Desarrollo de herramientas de minería de datos para HathiTrust, y uso de herramientas analíticas provenientes de otras fuentes**. HathiTrust ha participado en varias iniciativas estratégicas para dar soporte a la minería de datos en sus colecciones:
  - ***Distribución de datos***: HathiTrust ha puesto a disposición de los investigadores conjuntos de datos de ejemplo para el procesado y análisis computacional.
  - ***Centro de investigación***: HathiTrust ha creado un centro de investigación equipado con diversas herramientas y servicios para permitir realizar una gran variedad de análisis en su corpus de repositorios.

La biblioteca digital

La biblioteca digital de HathiTrust es un almacén de **preservación digital** y una plataforma de **acceso** altamente funcional. Proporciona servicios de preservación y acceso a largo plazo para contenido de dominio público y con derechos de autor de diversidad de fuentes, incluidos Google, el *Internet Archive*, Microsoft e iniciativas propias de instituciones asociadas.

Los socios garantizan la fiabilidad y eficiencia de la biblioteca digital apoyándose en los estándares y mejores prácticas de la comunidad, desarrollando políticas y procedimientos para la **gestión de contenidos y servicios** escalados, y manteniendo una infraestructura modular y abierta.

En la actualidad, hay digitalizados más de 14 millones de volúmenes, que representan aproximadamente unos 5000 millones de páginas y 636 terabytes. De todos ellos, aproximadamente el 39% son de dominio público.

HathiTrust proporciona una referencia para la implantación de **herramientas de acceso** al contenido del almacén, y trabaja con las bibliotecas participantes para definir, priorizar y desarrollar otras herramientas y servicios. También ha elaborado definiciones de servicio abierto (API) para hacer posible la participación de las bibliotecas asociadas en el desarrollo de otros mecanismos de acceso seguro:

- **Búsqueda bibliográfica**. La búsqueda bibliográfica (título, autor, asunto, ISBN, editor y año de publicación) está disponible a través de un cuadro de búsqueda del catálogo principal, en la página de inicio.
- **Búsqueda global**. HathiTrust ofrece además la posibilidad de realizar una búsqueda por texto completo para todos los elementos del almacén (incluidos trabajos de dominio público y con derechos de autor).
- **Visor y pasador de páginas** (*Page Turner*). HathiTrust ofrece un mecanismo para visualizar y pasar páginas de volúmenes individuales de dominio público. Es capaz de reproducir documentos PDF, texto de OCR (Reconocimiento Óptico de Caracteres, *Optical Character Recognition*) y archivos propios de un navegador.
- **Constructor de colecciones**. El constructor de colecciones ofrece a los usuarios finales y a los encargados del desarrollo de la colección la capacidad de crear y publicar colecciones virtuales o volúmenes mantenidos en el almacén.
- **Mecanismos de acceso para usuarios con discapacidades lectoras**. Se ha desarrollado un mecanismo para favorecer el acceso a la biblioteca a personas con discapacidades lectoras, cuyo uso pretende generalizarse para que funcione en toda institución asociada.
- **APIs**. HathiTrust distribuye información sobre elementos integradores de su almacén a través de una serie de mecanismos: Conjunto de datos, API bibliográficas y de datos, archivos delimitados por tabulador, archivos OCLC o fuentes de la Iniciativa de Archivos Abiertos (*Open Archives Initiative*, OAI).
- **Opciones de búsqueda adicionales**. Se ofrece a las instituciones asociadas la posibilidad de cargar registros bibliográficos pertenecientes al conjunto de materiales de HathiTrust en sus propios catálogos.

Políticas

Acceso y uso

HathiTrust es una iniciativa bibliotecaria colaborativa. Se anima a los usuarios a citar e incorporar vínculos a su contenido digital, y pueden hacerlo sin necesidad de pedir permiso. Dependiendo del origen del material digitalizado, las licencias y otros aspectos contractuales, se puede restringir su distribución a otros usuarios.

Corrección de metadatos bibliográficos

Se asume que la gestión de los metadatos bibliográficos que definen los registros bibliográficos de HathiTrust es asumida por los colaboradores de los diferentes catálogos. Por lo tanto, la política general consiste en no corregir o actualizar el contenido de los registros de los colaboradores, salvo cuando sea necesario a fin de garantizar la coordinación de las funciones del sistema de gestión de metadatos.

Derechos de autor

Siempre que sea posible, HathiTrust aboga por el dominio público. Sin embargo, hay muchos trabajos de sus colecciones que están protegidas por leyes de derechos de autor, de manera que no se puede mostrar grandes porciones de estas obras sin el permiso del detentor de los derechos de autor. Mientras no se pueda determinar el estatus de una obra, el acceso a la misma queda restringido.

Eliminación de HathiTrust

En muy raras ocasiones se producen eliminaciones en HathiTrust, y sólo ocurre cuando:

- El volumen es inservible debido a su falta de calidad, o bien existe una copia de calidad superior.
- La eliminación ha sido solicitada por el poseedor de los derechos de autor.

Preservación digital

HathiTrust se rige por los principios de fiabilidad, transparencia y gestión responsable. Proporciona una conservación garantizada a largo plazo del contenido digitalizado, así como un acceso abierto en la medida de lo legalmente posible, con el fin de maximizar las contribuciones de las instituciones asociadas y hacer un uso lo más eficiente posible de los recursos disponibles.

HathiTrust se compromete a preservar el contenido intelectual, y en muchos casos también el aspecto exacto de los materiales que han sido digitalizados para su depósito. Esto incluye:

- Representaciones digitales (imágenes) del contenido, tal y como aparece originalmente, con el mismo color y diseño (por ejemplo, para ilustraciones y obras artísticas), y en el mismo orden.
- Representaciones textuales de contenido, con tecnología OCR siempre que sea posible.

HathiTrust hace uso de diversas estrategias para asegurar la integridad a largo plazo de los materiales depositados, incluidas:

- Uso de formatos de contenido estándar y en abierto aceptados por la comunidad para la preservación digital, que sean soportados por multitud de plataformas y que se confíe que podrán ser preservados y migrados a nuevos formatos de preservación a lo largo del tiempo.
    - HathiTrust confía actualmente en las especificaciones de formatos de archivo, metadatos de preservación y métodos de control de calidad detallados en las especificaciones de digitalización de la Universidad de Míchigan.[4]
    - HathiTrust se compromete con la preservación a nivel de bit y la migración de formatos de los materiales creados de acuerdo con estas especificaciones y con la tecnología,

estándares y mejores prácticas que prevalecen en la comunidad bibliotecaria.
    - Entre los formatos preservados en HathiTrust se incluyen archivos TIFF con compresión ITU G4 almacenados a 600 ppp, archivos JPEG o JPEG2000 almacenados a diversas resoluciones entre 200 ppp y 400 ppp, texto Unicode, y archivos XML con un DTD asociado (típicamente METS, *Metadata Encoding and Transmission Standard*).

- Validación rigurosa del contenido incorporado. Confianza en estándares para el diseño del almacén y la confiabilidad tales como OAIS (*Open Archival Information System*) y TRAC.
- Confianza en estándares para metadatos como METS y PREMIS.
- Comprobaciones regulares de la integridad del contenido almacenado mediante:
    - Sistema de comprobación automática que verifica la integridad de los objetos digitales con sus versiones incorporadas. Esto se lleva a cabo trimestralmente, para todos los archivos.
    - Acceso de usuarios.
    - Procesos del almacén, como la indexación de texto completo haciendo uso del contenido de forma regular.

Así pues, la preservación en HathiTrust abarca características de contenido, metadatos y procesos que permiten mantener la integridad a nivel de bit del contenido a lo largo del tiempo, y migrar el contenido a nuevos formatos conforme lo requieran las necesidades en la comunidad bibliotecaria en cuestiones tecnológicas, estándares y mejores prácticas.

HathiTrust se esfuerza en garantizar que el contenido digital que preserva sea preciso, completo y adecuado para la conservación a largo plazo, además de útil para una gran diversidad de propósitos de acceso. Para ello presta atención a la calidad, y tiene en cuenta los formatos de los archivos de contenido, los metadatos de preservación y descriptivos y las rutinas de validación. HathiTrust mantiene un alto nivel de conformidad con los estándares de toda la comunidad de almacenes digitales, incluyendo el almacenamiento redundante de los materiales en puntos separados geográficamente.

Directrices de los depósitos digitales

El almacén de HathiTrust se diseñó de acuerdo con el marco para OAIS (*Open Archival Information Systems*), y ha sido implementado dentro del contexto de estándares y criterios ampliamente difundidos para repositorios digitales confiables *(Trustworthy Digital Repositories*). La logística de operación de un almacén de preservación de la dimensión de HathiTrust ha dado lugar a soluciones de implementación que favorecen la consistencia y estandarización frente a los cambios, la simplicidad frente a la complejidad (en diseño, no funcionalmente), y el aspecto práctico frente al conceptual. La funcionalidad de HathiTrust se consagra por encima de todo a la satisfacción de las necesidades de preservación y acceso de sus socios. Aunque HathiTrust sirva por extensión a un público más amplio, son estas necesidades específicas las que guían el desarrollo de los servicios y capacidades de HathiTrust.

Por otra parte, hay dos componentes para la incorporación y procesamiento en HathiTrust: metadatos bibliográficos y contenido.

Calidad

En todo caso, el material de HathiTrust está sujeto a revisiones de calidad, como parte integral y paso fundamental dentro del proceso de digitalización. El material digitalizado está sometido a procesos formales de revisión de la calidad, previamente a su entrada a formar parte del contenido digital de HathiTrust.

Privacidad

HathiTrust respeta la privacidad de todos los visitantes y usuarios de sus servicios.

Tecnología, estándares y certificaciones

Aspectos tecnológicos

HathiTrust proporciona almacenamiento persistente y con alta disponibilidad para los archivos depositados en su repositorio. Con el fin de facilitar esto, los socios emplean una arquitectura de almacenamiento con amplio abanico de funcionalidades diseñada para hacer frente a la tolerancia frente a fallos y la retención de datos a largo plazo.

La necesidad de una comprobación continua de la integridad es fundamental para la estrategia de gestión de datos de HathiTrust, y remarca la necesidad de elección de un medio principal en línea (discos magnéticos) La sustitución del material es un proceso que se contempla anualmente, y asume que el equipamiento tiene una vida útil de entre 3 y 4 años. El sistema de almacenamiento es modular y virtualizado, con archivos que se dividen en bloques distribuidos a través de los nodos de un clúster y redistribuidos de forma automática para realizar el balanceo de carga.

HathiTrust presenta un perfil de repositorio basado en la Evaluación de Sistemas de Publicación Electrónica de Origen Abierto[5] (*Evaluation of Open-Source Electronic Publishing Systems*) y un marco desarrollado específicamente *ad hoc*.

Estándares de la biblioteca y el contenido digital

HathiTrust está comprometido con la transparencia en todas sus operaciones, incluido su trabajo para cumplir con los estándares de preservación digital y los procesos de revisión. Representantes del Centro de Conservación Digital (https://web.archive.org/web/20140527162622/http://www.dcc.ac.uk/about-us) británico (*Digital Curation Centre*, DCC) y Preservación Digital en Europa (http://en.nationaalarchief.nl/information-management-and-creation-of-archives/sustainable-management-of-digital-archiva-23) (*Digital Preservation Europe*, DPE) revisaron el almacén usando el marco DRAMBORA a finales de 2008.

Además, HathiTrust cumple con otros estándares aceptados para la preservación digital, como los mostrados a continuación:

- **Auditoria y Certificación de Repositorios de Confianza** (*Trustworthy Repositories Audit & Certification*, TRAC): **Criterios y lista de verificación** (2007).
  El documento de TRAC fue desarrollado por el grupo de trabajo de certificación de almacenes digitales de RLG-NARA (*Research Libraries Group & National Archives and Records Administration Digital Repository Certification Task Force*), como resultado de una investigación sobre los medios para auditar y certificar archivos digitales, y se completaron una serie de pruebas para informar sobre la investigación. TRAC es el punto de origen para un proyecto de desarrollo de estándares ISO sobre auditoría y certificación de archivos digitales.
  HathiTrust ha efectuado una respuesta a los elementos requeridos en los criterios y lista de verificación del TRAC. El cumplimiento de TRAC del almacén fue certificado en 2011 por el Centro para Bibliotecas de Investigación (http://www.crl.edu/) norteamericano (*Center for Research Libraries*, CRL), y el informe de auditoría está disponible en su sitio web.
- **Sistemas de Información Archivística Abiertos** (*Open Archival Information System*, OAIS): **Modelo de referencia** (2002).

El modelo de referencia de OAIS fue desarrollado por el grupo de trabajo de un comité de sistemas de datos espaciales de la NASA y publicado en 2002. OAIS se aceptó como estándar ISO en 2003, y tuvo su continuación en 2007. OAIS conceptualiza y define las funciones, roles y contenido de un archivo digital.
HathiTrust se compromete a desarrollar sus propios sistemas y prácticas de preservación digital que cumplen con el estándar del modelo de referencia OAIS. Por ejemplo, lo relacionado con el Paquete de Presentación de la Información (*Submission Presentation Package*, SIP), que se concentra en el Paquete de Información de Archivo (*Archival Information Package*, AIP).

- **Estrategias de Implementación de Metadatos de Preservación** (*Preservation Metadata: Implementation Strategies*, PREMIS): **Diccionario** (2005).
  El diccionario de datos PREMIS y sus documentos asociados fueron desarrollados por un grupo de trabajo convocado por OCLC (*Online Computer Library Center*) y RLG (Grupo de Bibliotecas de Investigación, *Research Libraries Group*). Su mantenimiento y desarrollo está organizado por la Biblioteca del Congreso de EE. UU. (*Library of Congress*, LoC).
  HathiTrust documenta las fechas de los actos de preservación, así como las unidades semánticas obligatorias, en archivos METS para cada volumen, de acuerdo con el diccionario de datos PREMIS.

Especificaciones para los objetos digitales

El almacén de HathiTrust fue creado de acuerdo con el marco de Sistemas de Información Archivística Abiertos (*Open Archival Information Systems*, OAIS).

Definiciones

- Paquete de Información Archivística (*Archival Information Package*, AIP): El Paquete de Información, consistente en Información de Contenido y su Información de Descripción de Preservación (*Preservation Description Information*, PDI) asociada, que se conserva dentro de HathiTrust.
- Paquete de Información de Envío (*Submission Information Package*, SIP): El Paquete de Información entregada a HathiTrust y empleada para la construcción de una o más AIPs.
- Información de contenido: El conjunto de información que en última instancia se quiere conservar. Es un objeto de información que abarca al Objeto de Datos de Contenido y su Información de Representación.
  - Objeto de Datos de Contenido: el objeto de dato que, junto con la Información de Representación es originalmente objeto de preservación (en HathiTrust son en la actualidad archivos de imagen de página y archivos OCR y metadatos asociados).
  - Información de Representación: La información que mapea un Objeto de Datos en conceptos más significativos (incluye estándares a muy bajo nivel, como Unicode y TIFF).
- Información de Descripción de Preservación: La información necesaria para adecuar la preservación de la Información de Contenido y que puede categorizarse en Información de Procedencia, Referencia, Continuidad y Contexto.
  - Información de Procedencia (*Provenance*): Documenta la historia de la Información de Contenido, incluida su creación, cualquier alteración de su contenido o formato a lo largo del tiempo, su cadena de custodia, cualquier acción (como un refresco de los medios o una migración), efectuada para preservar la información del contenido, y el resultado de estas acciones.

- Información de Referencia (*Reference*): Identifica unívocamente la Información de Contenido en HathiTrust (por ejemplo, el identificador de repositorio), así como en relación con entidades y sistemas externos a HathiTrust (por ejemplo, número OCLC, ISBN, etc.).
- Información de Continuidad (*Fixity*): Valida la autenticidad o integridad de la Información de Contenido. Por ejemplo, un checksum, una firma digital o una marca de agua digital.

Especificaciones

La información de Procedencia, Referencia y Continuidad se almacena en HathiTrust en uno o más archivos que cumplen con el Estándar de Codificación y Transmisión de Metadatos (*Metadata Encoding and Transmission Standard*, METS). Los objetos digitales de los Paquetes de Información Archivística de todas las fuentes digitalizadas incluyen un archivo METS. Los que provienen del Internet Archive y de Google tienen un archivo METS «origen» adicional. Estos dos archivos se conforman de la siguiente manera:

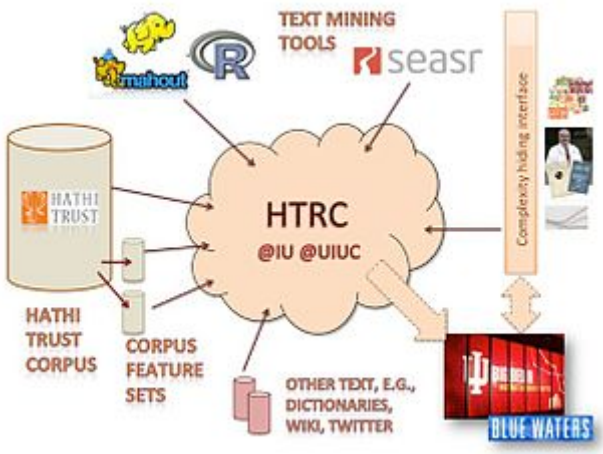
- Un archivo METS «origen» se ensambla con los metadatos proporcionados por HathiTrust en el Paquete de Información de Envío, y contiene información sobre la Información de Contenido desde el momento en que se creó hasta el instante en que entró en el almacén.
- Un archivo METS de HathiTrust se crea durante la incorporación, e incluye un subconjunto de datos del archivo METS «origen», pero fundamentalmente es un registro del objeto digital desde el momento en que se introduce en el almacén.

La información de preservación incluida en el archivo METS se registra utilizando las Estrategias de Implementación de Metadatos de Preservación (*Preservation Metadata Implementation Strategies*, PREMIS).

HathiTrust ha definido un perfil METS para el contenido digitalizado de Google almacenado en el archivo, y había definido una política general y un marco de especificaciones para contenido de libros y revistas (incluyendo los metadatos de encabezado de imagen, resolución, identificadores, etc.).

El centro de investigación

El Centro de Investigación de HathiTrust[6] (*HathiTrust Research Center*, HTRC) permite que usuarios de entidades educativas y organizaciones sin ánimo de lucro tengan acceso electrónico a trabajos publicados de dominio público disponibles en HathiTrust. Este acceso se extenderá en condiciones de uso restringido a trabajos con derechos de autor.



Centro de Investigación de HathiTrust (*HathiTrust Research Center*, HTRC).

Se trata de un centro de investigación colaborativo que se lanzó con el fin de satisfacer los desafíos técnicos a los que los investigadores se enfrentan al tratar con grandes cantidades de texto digital, mediante el desarrollo de herramientas de software de última generación y una infraestructura que permita el acceso electrónico avanzado al creciente archivo digital que abarca el conocimiento humano.

Así pues, el HTRC proporciona una infraestructura destinada a investigadores de entidades educativas y organizaciones sin ánimo de lucro para buscar, obtener, analizar y visualizar el texto completo a millones de obras de dominio público.

HTRC oculta la complejidad de la investigación computacional del corpus masivo de HathiTrust. Un investigador interacciona con HTRC a través de una interfaz que encapsula la funcionalidad y oculta la complejidad de su implementación.

La interfaz proporciona un portal Web y una interfaz de programación. HTRC reúne varias herramientas de minería de texto, el *corpus* de HathiTrust, información agregada y estadística sobre el corpus y otros orígenes de datos necesarios para la minería de texto.

Conclusiones

HathiTrust permite a los académicos acceder a un vasto abanico de materiales, producir búsquedas personalizadas, y descubrir nueva información que con anterioridad era difícilmente accesible. HathiTrust realza el valor de estos recursos garantizando el acceso a largo plazo, creando herramientas académicas y mejorando la calidad del contenido digital a lo largo del tiempo. Los investigadores se benefician así de la orientación de expertos y el acceso consistente del que siempre han hecho gala las bibliotecas de investigación, pero con la diferencia de que, en vez de tener que buscar uno a uno en los repositorios de cada institución, se benefician de una colección compartida. El todo es más que la suma de las partes.

Véase también

- Biblioteca digital
- Open Archival Information System
- Preservación digital

Referencias

1. «Significado de Hathi» (https://www.hathitrust.org/help_general#WhatDoesName) (en inglés).
2. «Socios de la comunidad de HathiTrust» (https://www.hathitrust.org/community).
3. Liene Karels (noviembre de 2010). «HathiTrust añade nuevos miembros y pasa a ser internacional» (http://ur.umich.edu/1011/Nov22_10/1807-hathitrust-adds-new). *The University Record Online (Universidad de Michigan)* (en inglés).
4. «Especificaciones de digitalización de la Universidad de Michigan» (http://www.lib.umich.edu/sites/default/files/services/dlps/umdigitizationspecs20140825-by-cc-license.pdf) (en inglés).
5. Cyzyk y Choudury (2008). «Un estudio y evaluación de los Sistemas de Publicación Electrónica Abiertos» (https://jscholarship.library.jhu.edu/handle/1774.2/32737). *JScholarship* (en inglés).
6. «Un estudio y evaluación de los Sistemas de Publicación Electrónica Abiertos» (https://www.youtube.com/watch?v=SVml1pDkMYk) (en inglés).

Enlaces externos

- Sitio web oficial (http://hathitrust.org/)
- HathiTrust incorpora nuevos miembros y pasa a ser internacional (http://ur.umich.edu/1011/Nov22_10/1807-hathitrust-adds-new/) (publicación oficial de prensa oficial del 13 de noviembre de 2010, en inglés)
- Los principales socios bibliotecarios lanzan el almacén digital compartido HathiTrust (https://www.hathitrust.org/press_10-13-2008/) (publicación oficial de prensa de 13 de octubre de 2008, en inglés)
- HathiTrust en la Universidad de Indiana (https://libraries.indiana.edu/hathitrust-digital-library/)

Obtenido de «https://es.wikipedia.org/w/index.php?title=HathiTrust&oldid=127407422»

Esta página se editó por última vez el 1 jul 2020 a las 22:28.

El texto está disponible bajo la Licencia Creative Commons Atribución Compartir Igual 3.0; pueden aplicarse cláusulas adicionales. Al usar este sitio, usted acepta nuestros términos de uso y nuestra política de privacidad.
Wikipedia® es una marca registrada de la Fundación Wikimedia, Inc., una organización sin ánimo de lucro.